MAS223 Statistical Modelling and Inference

Solutions to Semester 1 exercises

1. (a) $P(X = 3) = 1 - 0.4 - 0.4 = 0.2$
   (b) $E(X) = 1 \times 0.4 + 2 \times 0.4 + 3 \times 0.2 = 1.8.$
   (c) $E(X^2) = 1\times 0.4 + 4\times 0.4 + 9\times 0.2 = 3.8$, so $\text{Var}(X) = 3.8 - (1.8)^2 = 0.56.$

2. (a) $\int_{1/2}^{1} (y/2)\ dy = 3/16$
   (b) $F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(u)\ du = \int_0^y (u/2)\ du = y^2/4$ if $y \in [0, 2]$. If $y < 0$ $F(y) = 0$ and if $y > 2$ $F(y) = 1$.
   (c) $E(Y) = \int_0^2 y(y/2)\ dy = \int_0^2 (y^2/2)\ dy = 4/3$. $E(Y^2) = \int_0^2 (y^3/2)\ dy = 2$, so $\text{Var}(Y) = 2 - 16/9 = 2/9$.

3. (a) Binomial with parameters 8 and 1/6, or $Bi(8, 1/6)$.
   (b) $P(X = 2) = \binom{8}{2}(1/6)^2(5/6)^6 = 0.26$ (2 d.p.)

4. $\frac{df}{d\theta} = (4 - 2\theta)e^{-\theta^2+4\theta}$ and $\frac{d^2f}{d\theta^2} = ((4 - 2\theta)^2 - 2)e^{-\theta^2+4\theta}$. $\frac{df}{d\theta}$ is zero only when $\theta = 2$, and at $\theta = 2$ $\frac{d^2f}{d\theta^2} < 0$, so this is a maximum. There are no other turning points or discontinuities, so it is the global maximum.

5. (a) $\int_{x=0}^{1}\int_{y=0}^{1} f(x, y)\ dy\ dx = \int_{x=0}^{1}(e^{-x} - e^{-(1+x)})\ dx = e^0 - e^{-1} - e^{-1} + e^{-2} = 1 + e^{-2} - 2e^{-1}$;
   (b) $\int_{x=0}^{\infty} f(x, y)\ dx = [-e^{-(x+y)}]_0^{\infty} = e^{-y} - 0 = e^{-y}$

6.
$$\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 15 & 0 \\ 0 & 21 \end{pmatrix}.$$

7. $E(T_1) = E(T_2) = E(T_3) = \mu$ (so all unbiased). $\text{Var}(T_1) = \frac{5}{4}$, $\text{Var}(T_2) = 8$, $\text{Var}(T_3) = \frac{4}{5}$. On this information, prefer T_3 as smallest mean square error.

8. (a) Using partial fractions we have

$$\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$$

and so for any $x = 1, 2, \ldots$

$$\begin{aligned} F(x) &= \sum_{u=1}^{x} p(u) = \sum_{u=1}^{x} \frac{1}{u(u+1)} = \sum_{u=1}^{x} \left(\frac{1}{u} - \frac{1}{u+1} \right) = \sum_{u=1}^{x} \frac{1}{u} - \sum_{u=1}^{x} \frac{1}{u+1} \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{x} \\ &\quad - \frac{1}{2} - \frac{1}{3} - \cdots - \frac{1}{x+1} \\ &= 1 - \frac{1}{x+1} \end{aligned}$$

So the d.f is

$$F(u) = \begin{cases} 1 - \frac{1}{x+1}, & \text{for } 1 \le x \le u < x+1, x \in \mathbb{N} \\ 0, & \text{for } u < 1 \end{cases}$$

See the figure for roughly what the graph should look like.

(b)

$$\begin{aligned} P(10 \le X \le 20) &= P(X \le 20) - P(X \le 9) \\ &= F(20) - F(9) \\ &= 1 - \frac{1}{21} - 1 + \frac{1}{10} \\ &= \frac{11}{210} \text{ or } 0.052. \end{aligned}$$

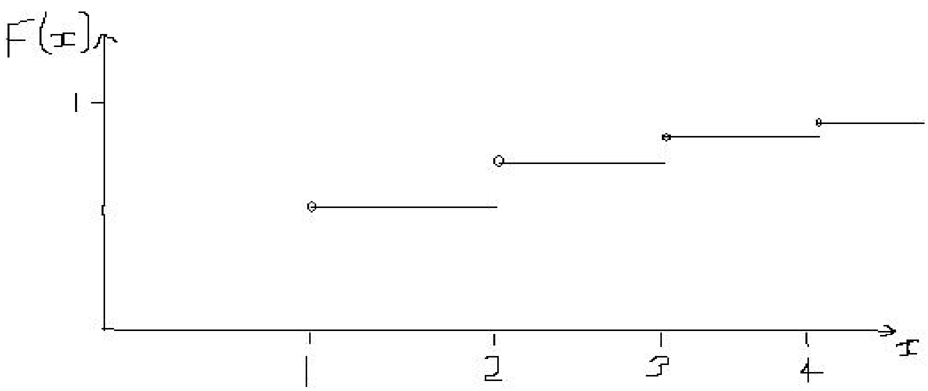


Figure 1: Sketch graph for question 8(a)

(c)

$$\begin{aligned} 0.01 \geq P(X \geq x) &= 1 - P(X < x) \\ &= 1 - P(X \leq x - 1) \end{aligned}$$

so $P(X \leq x-1) \geq 0.99$ and hence

$$F(x-1) \geq 0.99. \tag{1}$$

Now from the d.f $F(x)$ we have

$$0.99 = 1 - \frac{1}{x+1} \Rightarrow \frac{1}{x+1} = 0.01 \Rightarrow x = 99$$

and so by (1)

$$F(x-1) \geq 0.99 = F(99) \Rightarrow x - 1 \geq 99 \Rightarrow x \geq 100$$

and so the required minimum number is 100.

(d) Attempting to calculate $E(X)$ gives

$$\sum_{x=1}^{\infty} xp(x) = \sum_{x=1}^{\infty} \frac{1}{x+1},$$

which is an infinite sum. So $E(X)$ is not defined.

9. (a) F is a non-decreasing function since

$$F'(x) = \frac{e^x(1+e^x) - e^x e^x}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2} > 0.$$

(In fact F is a strictly increasing function). Then we have

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \frac{e^x}{1+e^x} = \lim_{x\to-\infty} \frac{e^x}{e^x(e^{-x}+1)} = \lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0$$

and

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} \frac{e^x}{1+e^x} = \lim_{x\to\infty} \frac{e^x}{e^x(e^{-x}+1)} = \lim_{x\to\infty} \frac{1}{1+e^{-x}} = 1.$$

Thus all requirements of the definition of the distribution function are satisfied and so $F(x)$ is a continuous distribution function.

(b) From part (a)

$$f(x) = F'(x) = \frac{e^x}{(1+e^x)^2}.$$

Now

$$\begin{aligned} f(-x) = \frac{e^{-x}}{(1+e^{-x})^2} &= \frac{1}{e^x} \frac{1}{\left(1+\frac{1}{e^x}\right)^2} \\ &= \frac{1}{e^x} \frac{1}{\left(\frac{e^x+1}{e^x}\right)^2} \\ &= \frac{e^x}{(1+e^x)^2} \\ &= f(x). \end{aligned}$$

(c)

$$\begin{aligned}
P(|X| > 2) &= P(X > 2 \quad \text{or} \quad X < -2) = P(X > 2) + P(X < -2) \\
&= 1 - P(X \le 2) + P(X < -2) = 1 - F(2) + F(-2) \\
&= 1 - \frac{e^2}{1+e^2} + \frac{e^{-2}}{1+e^{-2}} = 0.238.
\end{aligned}$$

10. (a) $F(x)$ does not converge to 1 as $x \to \infty$.
(b) $x + \frac{1}{4}\sin(2\pi x)$ is not a non-decreasing function. (You can see this by noticing that its derivative is negative around $x = 1/2$.)

11. Let $-1 \le x < 0$, then $F(-1) = P(X \le -1) = P(X < -1) = 0$ and

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(u)\,du \\
&= F(-1) + \int_{-1}^{x} (1+u)\,du \\
&= \left[u + \frac{u^2}{2}\right]_{-1}^{x} \\
&= x + \frac{x^2}{2} + 1 - \frac{1}{2} \\
&= \frac{(x+1)^2}{2}.
\end{aligned}$$

Now for $0 \le x < 1$, first we calculate $F(0) = P(X \le 0) = P(X < 0) = \frac{1}{2}$

and

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(u)\,du \\
&= F(0) + \int_{0}^{x} (1-u)\,du \\
&= \frac{1}{2} + \left[u - \frac{u^2}{2}\right]_0^x \\
&= \frac{1}{2} + x - \frac{x^2}{2} \\
&= \frac{1+2x-x^2}{2}.
\end{aligned}
$$

Thus the distribution function $F(x)$ is

$$
F(x) = \begin{cases} 0, & \text{for } x < -1 \\ \frac{(x+1)^2}{2}, & \text{for } -1 \leq x < 0 \\ \frac{1+2x-x^2}{2}, & \text{for } 0 \leq x < 1 \\ 1, & \text{for } x \geq 1 \end{cases}
$$

12. For X, the mean μ is 1.8 (from the solutions to Exercise 1), so $E((X-\mu)^3) = 0.4 \times (-0.8)^3 + 0.4 \times (0.2)^3 + 0.2 \times (1.2)^3 = 0.144$, and $\sigma^2 = 0.56$, so the coefficient of skewness is

$$
\frac{E((X-\mu)^3)}{\sigma^3} = \frac{0.144}{\left(\sqrt{0.56}\right)^3} = 0.344.
$$

For Y, $\mu = 4/3$ and $\sigma^2 = 2/9$, so

$$\begin{aligned} E((Y-\mu)^3) &= \int_0^2 \frac{y(y-4/3)^3}{2}\,dy \\ &= \int_0^2 \left(\frac{y^4}{2} - 2y^3 + \frac{8y^2}{3} - \frac{32y}{27}\right)\,dy \\ &= -\frac{8}{135}, \end{aligned}$$

so the coefficient of skewness is

$$\frac{E((Y-\mu)^3)}{\sigma^3} = \frac{-8/135}{\left(\sqrt{2/9}\right)^3} = -0.566.$$

13. From Example 3, with $N = 10$, $N^* = 7$ and $n = 5$, we have

$$p(x) = P(X = x) = \frac{\binom{7}{x}\binom{3}{5-x}}{\binom{10}{5}}$$

where x satisfies $x \geq 0$, $x \leq N^* = 7$, $n - x \geq 0 \Rightarrow x \leq 5$, $n - x \leq N - N^* \Rightarrow x \geq 2$. All these inequalities imply $2 \leq x \leq 5$, so the range $R_X = \{2, 3, 4, 5\}$. Thus

$$p(2) = \frac{\binom{7}{2}\binom{3}{3}}{\binom{10}{5}} = \frac{21}{252} = \frac{1}{12} \approx 0.083$$

$$p(3) = \frac{\binom{7}{3}\binom{3}{2}}{\binom{10}{5}} = \frac{105}{252} = \frac{5}{12} \approx 0.416$$

$$p(4) = \frac{\binom{7}{4}\binom{3}{1}}{\binom{10}{5}} = \frac{105}{252} = \frac{5}{12} \approx 0.416$$

$$p(5) = \frac{\binom{7}{5}\binom{3}{0}}{\binom{10}{5}} = \frac{21}{252} = \frac{1}{12} \approx 0.083$$

You can check: $\sum_{x \in R_X} p(x) = p(2) + p(3) + p(4) + p(5) = 1.$

For the mean of X we have

$$E(X) = \sum_{x \in R_X} xp(x) = 2p(2) + 3p(3) + 4p(4) + 5p(5) = \frac{42 + 315 + 420 + 105}{252} = 3.5$$

For the variance of X, first we calculate $E(X^2)$,

$$E(X^2) = \sum_{x \in R_X} x^2 p(x) = 4p(2) + 9p(3) + 16p(4) + 25p(5) = 12.833.$$

So the variance of X is

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 12.807 - 3.5^2 = 0.583.$$

14. (a) First we find the d.f. of X.

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \lambda \int_{0}^{x} e^{-\lambda u}\,du = \lambda \left[-\frac{1}{\lambda} e^{-\lambda u} \right]_0^x = 1 - e^{-\lambda x}.$$

The hazard function $h(t)$ is

$$h(t) = \frac{\lambda e^{-\lambda t}}{1 - 1 + e^{-\lambda t}} = \lambda \quad \text{(constant in the age } t\text{)}$$

(b) The p.d.f. of X is

$$f(t) = F'(t) = \frac{1+t-t}{(1+t)^2} = \frac{1}{(1+t)^2}$$

and the hazard function is

$$h(t) = \frac{\frac{1}{(1+t)^2}}{1-\frac{t}{1+t}} = \frac{1}{1+t} \quad \text{(decreasing in the age } t\text{)}$$

(c) The p.d.f. is

$$f(t) = F'(t) = 2te^{-t^2}$$

and so the hazard function is

$$h(t) = \frac{2te^{-t^2}}{1-1+e^{-t^2}} = 2t \quad \text{(increasing in the age } t\text{)}$$

15. From the definition of the p.d.f. of the gamma distribution (see Example 6) we have

$$\begin{aligned}
E(X^k) &= \int_0^\infty x^k \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{k+\alpha-1} e^{-\beta x}\, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(k+\alpha)}{\beta^{k+\alpha}} \\
&= \frac{\Gamma(k+\alpha)}{\beta^k \Gamma(\alpha)} \\
&= \frac{(k+\alpha-1)(k+\alpha-2)\cdots(k+\alpha-k+1)(k+\alpha-k)\Gamma(\alpha)}{\beta^k\Gamma(\alpha)} \\
&= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}.
\end{aligned}$$

If $k = 1$ $E(X) = a/b$, which agrees with Example 6. If $k = 2$

$$E(X^2) = \frac{\alpha(\alpha+1)}{\beta^2}$$

and so the variance of X is

$$\mathrm{var}(X) = E(X^2) - [E(X)]^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2},$$

which, again, agrees with Example 6.

With $\mu = E(X)$ and $\sigma^2 = \mathrm{var}(X)$, the coefficient of skewness of X is

$$\begin{aligned}
\beta_1 &= \frac{E(X-\mu)^3}{\sigma^3} = \frac{E(X^3 - 3X^2\mu + 3X\mu^2 - \mu^3)}{\sigma^3} \\
&= \frac{E(X^3) - 3E(X)E(X^2) + 2[E(X)]^3}{\sigma^3} = \frac{\frac{\alpha(\alpha+1)(\alpha+2)}{\beta^3} - \frac{3\alpha^2(\alpha+1)}{\beta^3} + \frac{2\alpha^3}{\beta^3}}{\sigma^3} \\
&= \frac{\alpha(\alpha^2 + 2\alpha + \alpha + 2 - 3\alpha^2 - 3\alpha + 2\alpha^2)/\beta^3}{\frac{\sqrt{\alpha^3}}{\beta^3}} = \frac{2\alpha}{\sqrt{\alpha^3}} = \frac{2}{\sqrt{\alpha}}.
\end{aligned}$$

16. (a) We have $X \sim Be(\alpha, \beta)$ and $Y = c/X$ so that $y = c/x$ and $x = c/y$, which implies that

$$\left|\frac{dx}{dy}\right| = \frac{c}{y^2}.$$

Denote with $f_X(x)$ the p.d.f. of X and with $f_Y(y)$ the p.d.f. of Y. We have

$$\begin{aligned}
f_Y(y) &= f_X(y)\left|\frac{dx}{dy}\right| = \frac{1}{B(\alpha,\beta)}\left(\frac{c}{y}\right)^{\alpha-1}\left(1-\frac{c}{y}\right)^{\beta-1}\frac{c}{y^2} \\
&= \frac{c^\alpha}{B(\alpha,\beta)}\frac{(y-c)^{\beta-1}}{y^{\alpha+\beta}} \text{ for } y > c.
\end{aligned}$$

(b) There are a number of different ways of approaching this.

Method 1

(This seems to be the method that most of you actually attempted.)
Using the result of (a),

$$\begin{aligned} E(Y) &= \int_c^\infty y \cdot \frac{c^\alpha}{B(\alpha,\beta)} \frac{(y-c)^{\beta-1}}{y^{\alpha+\beta}}\, dy \\ &= \frac{1}{B(\alpha,\beta)} \int_c^\infty \left(\frac{c}{y}\right)^\alpha \left(1-\frac{c}{y}\right)^{\beta-1} dy. \end{aligned}$$

To make this look more like a Beta integral, change variables $u = c/y$. Because $\frac{dy}{du} = -c/u^2$ and because $y = c$ corresponds to $u = 1$ and $y = \infty$ corresponds to $u = 0$, we get

$$\frac{1}{B(\alpha,\beta)} \int_1^0 u^\alpha (1-u)^{\beta-1} \left(-\frac{c}{u^2}\right) du,$$

and reversing the order of the limits the minus sign disappears, giving

$$\frac{1}{B(\alpha,\beta)} \int_0^1 u^\alpha (1-u)^{\beta-1} \left(\frac{c}{u^2}\right) du = \frac{c}{B(\alpha,\beta)} \int_0^1 u^{\alpha-2}(1-u)^{\beta-1}\, du,$$

which is

$$\frac{c}{B(\alpha,\beta)} B(\alpha-1,\beta),$$

by the definition of the Beta function, as long as $\alpha > 1$ (as the Beta function requires both arguments to be positive).

[NB: the complications here over the minus sign arise because we've used a decreasing function in the change of variables. The key point is that the first integral after the change of variables is "upside down", with the lower limit greater than the upper one. Correcting this by swapping the limits cancels out the minus sign.]

Expanding the Beta functions here in terms of the Gamma function, we get

$$c\frac{\frac{\Gamma(\alpha-1)\Gamma(\beta)}{\Gamma(\alpha+\beta-1)}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} = c\frac{\frac{\Gamma(\alpha-1)\Gamma(\beta)}{\Gamma(\alpha+\beta-1)}}{\frac{(\alpha-1)\Gamma(\alpha-1)\Gamma(\beta)}{(\alpha+\beta-1)\Gamma(\alpha+\beta-1)}} = \frac{c(\alpha+\beta-1)}{\alpha-1}.$$

Method 2

(This method uses a trick which is often useful: we know that the integral of a p.d.f. must be 1, so if we can recognise a p.d.f. in our integral, perhaps with different parameters, we can often avoid doing any more integration by using what we already know.)

From (a) and from $\int_{-\infty}^{\infty} f_Y(y)\,dy = 1$ it follows that

$$\int_c^\infty \frac{(y-c)^{\beta-1}}{y^{\alpha+\beta}}\,dy = \frac{B(\alpha,\beta)}{c^\alpha}. \tag{2}$$

Now the mean of Y is

$$\begin{aligned} E(Y) &= \int_c^\infty y\cdot\frac{c^\alpha}{B(\alpha,\beta)}\frac{(y-c)^{\beta-1}}{y^{\alpha+\beta}}\,dy \\ &= \frac{c^\alpha}{B(\alpha,\beta)}\int_c^\infty \frac{(y-c)^{\beta-1}}{y^{\alpha+\beta-1}}\,dy \\ &= \frac{c^\alpha}{B(\alpha,\beta)}\frac{B(\alpha-1,\beta)}{c^{\alpha-1}} \\ &\quad \text{(using (2) with } \alpha-1 \text{ instead of } \alpha) \\ &= \frac{cB(\alpha-1,\beta)}{B(\alpha,\beta)}, \end{aligned}$$

from where we proceed as in Method 1.

Method 3

(This is perhaps the easiest.)

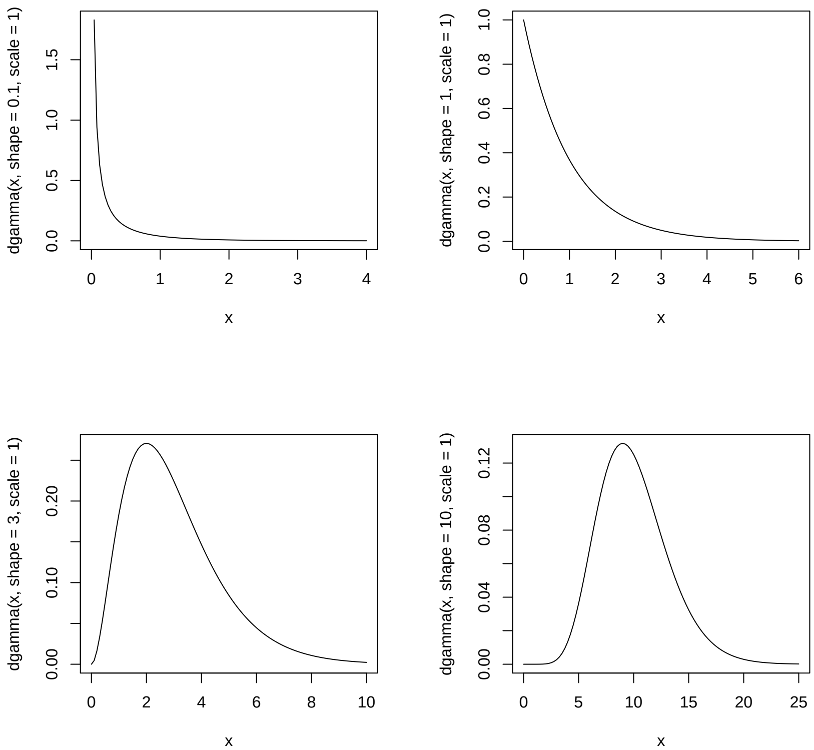
We can use the formula
$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$ (see §1.3 of the notes). So

$$\begin{aligned}
E\left(\frac{c}{X}\right) &= \int_{-\infty}^{\infty} \frac{c}{x} f_X(x)\,dx \\
&= \int_0^1 \frac{c}{x} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{c}{B(\alpha,\beta)} \int_0^1 x^{\alpha-2}(1-x)^{\beta-1}\,dx \\
&= \frac{cB(\alpha-1,\beta)}{B(\alpha,\beta)},
\end{aligned}$$

from where we proceed as in Method 1.

17. The following examples are with $\beta = 1$, and $\alpha = 0.1, 1, 3, 10$. For example the first picture was produced with

```
curve(dgamma(x,shape=0.1,scale=1),from=0,to=4)
```



The skewness appears to decrease with α. This is consistent with the value $2/\sqrt{\alpha}$ for the coefficient of skewness calculated in question 15.

18. (a) The function $g(x) = 5x + 3$ is increasing, so we can use the formula

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

We note $g^{-1}(y) = \frac{y-3}{5}$ and $\frac{d}{dy}g^{-1}(y) = \frac{1}{5}$, so

$$\begin{aligned} f_Y(y) &= \begin{cases} \left(1+\frac{y-3}{5}\right)\frac{1}{5} & -1 < \frac{y-3}{5} < 0 \\ \left(1-\frac{y-3}{5}\right)\frac{1}{5} & 0 < \frac{y-3}{5} < 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{y+2}{25} & -2 < y < 3 \\ \frac{8-y}{25} & 3 < y < 8 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(b) The function $g(x) = |x|$ is not monotonic, so we find

$$\begin{aligned} F_Z(z) &= P(Z \le z) \\ &= P(-z \le X \le z) \\ &= \begin{cases} \int_{-z}^{0}(1+u)\,du + \int_0^z (1-u)\,du & z \le 1 \\ 1 & z > 1 \end{cases} \\ &= \begin{cases} 2z - z^2 & z \le 1 \\ 1 & z > 1 \end{cases} \end{aligned}$$

So the p.d.f. of Z is

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \begin{cases} 2(1-z) & 0 \le z < 1 \\ 0 & \text{otherwise.} \end{cases}$$

19. The function $g(x) = \sqrt[r]{x}$ is increasing on $[0,1]$, so we can use

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y).$$

We have $f_X(x) = x^{\alpha-1}/B(\alpha,1) = \alpha x^{\alpha-1}$, $g^{-1}(y) = y^r$ and $\frac{d}{dy}g^{-1}(y) = ry^{r-1}$, so

$$f_Y(y) = \alpha y^{r(\alpha-1)} \cdot ry^{r-1} = r\alpha y^{r\alpha-1},$$

which is the p.d.f. of a $Be(r\alpha, 1)$ distribution. So Y has a Beta distribution with parameters $r\alpha$ and 1.

20. First of all, Θ has a $U(-\pi/2, \pi/2)$ distribution so, from the continuous distributions handout,

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases}$$

The function $g(\theta) = \tan\theta$ is increasing on the range of Θ, so we can use

$$\begin{aligned} f_X(x) &= f_\Theta(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x) \\ &= \frac{1}{\pi} \frac{1}{1+x^2}, \end{aligned}$$

using the derivative of $g^{-1}(x) = \tan^{-1}(x)$. Hence X has a t_1 (Cauchy) distribution.

21. (a) We need to integrate the joint p.d.f. over the subset of S where $u + v \leq 1$. This subset can be written $\{(u, v) : 0 \leq u \leq 1, 0 \leq v \leq 1 - u\}$, so we get

$$\begin{aligned} \int_0^1 \int_0^{1-u} \frac{4u + 2v}{3} \, dv \, du &= \frac{1}{3} \int_0^1 (4u(1-u) + (1-u)^2) \, du \\ &= \frac{1}{3} \int_0^1 (1 + 2u - 3u^2) \, du \\ &= \frac{1}{3}. \end{aligned}$$

(b) Integrate v out, giving (for $u \in [0, 1]$)

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, v) \, dv = \int_0^1 \frac{4u + 2v}{3} \, dv = \frac{4u + 1}{3}.$$

So the p.d.f. of U is

$$f_U(u) = \begin{cases} \frac{4u+1}{3} & 0 \leq u \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Check that this is a p.d.f.: it is non-negative and

$$\int_0^1 \frac{4u+1}{3}\, du = \left[\frac{2u^2+u}{3}\right]_0^1 = 1$$

as required.

(c) Integrate u out, giving (for $v \in [0,1]$)

$$f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\, du = \int_0^1 \frac{4u+2v}{3}\, du = \frac{2v+2}{3}.$$

So the p.d.f. of V is

$$f_V(v) = \begin{cases} \frac{2v+2}{3} & 0 \le v \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Check that this is a p.d.f.: it is non-negative and

$$\int_0^1 \frac{2v+2}{3}\, dv = \left[\frac{2v^2+2v}{3}\right]_0^1 = 1$$

as required.

22. (a) We need $\int_0^\infty \int_0^x f_{X,Y}(x,y)\, dy\, dx = 1$. So we calculate

$$\begin{aligned} k\int_0^\infty \int_0^x e^{-(x+y)}\, dy\, dx &= k\int_0^\infty \left[-e^{-(x+y)}\right]_{y=0}^x\, dx \\ &= k\int_0^\infty (e^{-x} - e^{-2x})\, dx \\ &= \frac{k}{2}, \end{aligned}$$

so $k = 2$.

(b) Integrate x out, giving, for $y > 0$,

$$f_Y(y) = 2\int_y^\infty e^{-(x+y)}\, dx = 2\left[-e^{-(x+y)}\right]_{x=y}^\infty = 2e^{-2y}.$$

So Y has an Exponential distribution with parameter 2 (or Gamma with parameters $\alpha = 1$ and $\beta = 2$).

23. (a) We need

$$f_{U|V}(u, v) = \frac{f_{U,V}(u, v)}{f_V(v)} = \frac{(4u + 2v)/3}{(2 + 2v)/3} = \frac{4u + 2v}{2 + 2v},$$

for $u \in [0, 1]$. So the conditional p.d.f. of U given $V = v$ is

$$f_{U|V}(u|v) = \begin{cases} \frac{2u+v}{1+v} & 0 \leq u \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(b) To find the conditional expectation, calculate

$$\int_0^1 u f_{U|V}(u|v)\, du = \frac{1}{1+v} \int_0^1 (2u^2 + uv)\, du = \frac{4 + 3v}{6(1 + v)}.$$

So $g(v) = \frac{4+3v}{6(1+v)}$

24. The marginal p.d.f. of X is given by $\int_0^1 f_{X,Y}(x, y)\, dy$. If $x \in [-1, 0]$ then this is $\int_0^1 \frac{y-x}{2}\, dy = \frac{1-2x}{4}$, and if $x \in [0, 1]$ it is $\int_0^1 \frac{x+y}{2}\, dy = \frac{1+2x}{4}$; otherwise it is zero. This gives $E(X) = 0$.

To find $E(XY)$ we calculate

$$\begin{aligned} \int_0^1 \int_{-1}^1 xy f_{X,Y}(x, y)\, dx\, dy &= \int_0^1 \int_{-1}^0 \frac{y^2x - x^2y}{2}\, dx\, dy + \int_0^1 \int_0^1 \frac{x^2y + y^2x}{2}\, dx\, dy \\ &= \int_0^1 \left(-\frac{y^2}{4} - \frac{y}{6} \right) dy + \int_0^1 \left(\frac{y^2}{4} + \frac{y}{6} \right) dy \\ &= 0. \end{aligned}$$

So the covariance $E(XY) - E(X)E(Y) = 0$ (note that we do not need the actual value of $E(Y)$). Hence the correlation coefficient is also zero.

The joint p.d.f. of X and Y does not factorise into marginal distributions for X and Y (dividing $f_{X,Y}(x,y)$ by $f_X(x)$ gives $\frac{2(y+x)}{1+2x}$ if $x \in [0,1]$ and $\frac{2(y-x)}{1-2x}$ if $x \in [-1,0]$, which does not depend only on y), so X and Y are not independent.

25. Using the properties of conditional expectation and variance in the notes and the mean and variance of Gamma and Normal random variables,

$$E(C) = E(E(C|S)) = E(kS) = k\frac{\alpha}{\beta}$$

and

$$\begin{aligned} \mathrm{Var}(C) &= E(\mathrm{Var}(C|S)) + \mathrm{Var}(E(C|S)) \\ &= E(\sigma^2) + \mathrm{Var}(kS) \\ &= \sigma^2 + k^2\frac{\alpha}{\beta^2}. \end{aligned}$$

So the mean is $\frac{k\alpha}{\beta}$ and the variance is $\sigma^2 + \frac{k^2\alpha}{\beta^2}$.

26. Define $U = X + Y$ and $W = X$ so that $Y = U - X = U - W$. From $y = u - w \geq 0$ we have $u \geq w$ and of course $u, w \geq 0$. Now

$$\frac{\partial x}{\partial u} = 0, \quad \frac{\partial x}{\partial w} = 1, \quad \frac{\partial y}{\partial u} = 1, \quad \frac{\partial y}{\partial w} = -1$$

and so the Jacobian is

$$\begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1.$$

We have

$$f_{U,W}(u,w) = f_{X,Y}(w, u-w) \times 1 = \begin{cases} \frac{1}{2}ue^{-u}, & u \geq 0, \quad 0 \leq w \leq u \\ 0, & \text{otherwise} \end{cases}$$

The p.d.f. of U is

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,W}(u,w)\,dw = \int_0^u \frac{1}{2}ue^{-u}\,dw = \frac{1}{2}ue^{-u}[w]_0^u = \frac{1}{2}u^2e^{-u} = \frac{1}{\Gamma(3)}u^{3-1}e^{-u}$$

and so $U \sim Ga(3,1)$. From Question 15, we have

$$E[(X+Y)^k] = E(U^k) = \frac{3(3+1)\cdots(3+k-1)}{1^k} = 3(4)\cdots(2+k)$$

which, for $k=5$, gives

$$E[(X+Y)^5] = 3(4)\cdots(7) = \frac{7!}{2} = 2520.$$

27. (a) We have

$$f_{X,Y} = \begin{cases} 2e^{-(x+y)} & x > y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and if $(u,v) = (x-y, y/2)$ then $(x,y) = (u+2v, 2v)$, with Jacobian

$$\begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} = 2.$$

So

$$f_{U,V}(u,v) = 4e^{-(u+4v)}$$

if $u+2v > 2v > 0$, i.e. if both u and v are positive, and zero otherwise.

(b) $f_{U,V}(u,v)$ factorises as $g(u)h(v)$ where

$$g(u) = \begin{cases} e^{-u} & u > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$h(v) = \begin{cases} 4e^{-4v} & v > 0 \\ 0 & \text{otherwise,} \end{cases}$$

so U and V are independent.

(c) We recognise $g(u)$ and $h(v)$ as the p.d.f.s of $Exp(1)$ and $Exp(4)$ random variables respectively, so $U \sim Exp(1)$ and $V \sim Exp(4)$.

28. Since X and Y are independent we have

$$
\begin{aligned}
f_{X,Y}(x,y) &= f_X(x)f_Y(y) = \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)}x^{\alpha_1-1}e^{-\beta x}\frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)}y^{\alpha_2-1}e^{-\beta y} \\
&= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}x^{\alpha_1-1}y^{\alpha_2-1}e^{-\beta(x+y)}. \qquad (3)
\end{aligned}
$$

If $v = x + y$ and $u = \frac{x}{x+y}$ then, to find the inverse transformation we observe that $uv = (x+y)\frac{x}{x+y} = x$, so $x = uv$ and hence $y = v - uv$. The conditions $x > 0$ and $y > 0$ translate to $v > 0$ and $0 < u < 1$. The partial derivatives are

$$
\frac{\partial x}{\partial u} = v, \quad \frac{\partial x}{\partial v} = u, \quad \frac{\partial y}{\partial u} = -v, \quad \frac{\partial y}{\partial v} = 1 - u
$$

and the Jacobian is

$$
\begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v(1-u) + uv = v - uv + uv = v > 0.
$$

Since $x, y \geq 0$, $v, u \geq 0$ and $u \leq 1$. From (1) The joint p.d.f. of U and V is

$$
\begin{aligned}
f_{U,V}(u,v) &= f_{X,Y}(uv, v-uv)v = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}(uv)^{\alpha_1-1}(v-uv)^{\alpha_2-1}e^{-\beta v}v \\
&= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}u^{\alpha_1-1}(1-u)^{\alpha_2-1}v^{\alpha_1+\alpha_2-1}e^{-\beta v},
\end{aligned}
$$

for $v \geq 0, 0 \leq u \leq 1$.

At this point we can observe that $f_{U,V}(u,v)$ factorises as

$$
f_{U,V}(u,v) = \frac{1}{B(\alpha_1,\alpha_2)}u^{\alpha_1-1}(1-u)^{\alpha_2-1}\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}v^{\alpha_1+\alpha_2-1}e^{-\beta v},
$$

(recall the relationship between the Beta and Gamma functions) showing that $U \sim Be(\alpha_1, \alpha_2)$ and $V \sim Ga(\alpha_1 + \alpha_2, \beta)$ and that U and V are independent.

Alternatively, we can calculate the marginal p.d.f.s. The marginal p.d.f. of U is

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} f_{U,V}(u,v)\,dv = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1-1}(1-u)^{\alpha_2-1}\int_0^{\infty} v^{\alpha_1+\alpha_2-1}e^{-\beta v}\,dv \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1-1}(1-u)^{\alpha_2-1}\frac{\Gamma(\alpha_1+\alpha_2)}{\beta^{\alpha_1+\alpha_2}} = \frac{1}{B(\alpha_1,\alpha_2)}u^{\alpha_1-1}(1-u)^{\alpha_2-1} \end{aligned}$$

and so $U \sim Be(\alpha_1, \alpha_2)$.

The marginal p.d.f. of V is

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_{UV}(u,v)\,du = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1+\alpha_2-1}e^{-\beta v}\int_0^{1} u^{\alpha_1-1}(1-u)^{\alpha_2-1}\,du \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1+\alpha_2-1}e^{-\beta v}\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}v^{\alpha_1+\alpha_2-1}e^{-\beta v} \end{aligned}$$

and so $V \sim Ga(\alpha_1+\alpha_2, \beta)$.

The independence of U and V follows by noticing that

$$f_{U,V}(u,v) = f_U(u)f_V(v),$$

the verification of which is left to the reader as a simple exercise.

29. (a) The case $n = 2$ is the given result from Question 28. If the claim is true for $n = k$, then let $X = \sum_{i=1}^{k} X_i$ and $Y \sim X_{k+1}$, so that $X \sim Ga\left(\sum_{i=1}^{k}\alpha_i, \beta\right)$ and $Y \sim Ga(\alpha_{k+1}, \beta)$. Then the same result from Question 28 shows that

$$\sum_{i=1}^{k+1} X_i = X + Y \sim Ga\left(\sum_{i=1}^{k+1}\alpha_i, \beta\right),$$

so the claim is true for $n = k+1$. Hence it is true for all $n \geq 2$ by induction.

(b) The χ_n^2 distribution is the same as $Ga\left(\frac{n}{2}, \frac{1}{2}\right)$. In particular each $Z_i^2 \sim Ga\left(\frac{1}{2}, \frac{1}{2}\right)$, so the result of (a) tells us that $\sum_{i=1}^n Z_i^2 \sim Ga\left(\frac{n}{2}, \frac{1}{2}\right) = \chi_n^2$ as required.

30. (a) We have $E((X,Y,Z)^T) = (E(X), E(Y), E(Z))^T = (3, -4, 6)^T$. Also $\text{Cov}(X,Y) = 0$ since X and Y are uncorrelated.
We have $\rho_{X,Z} = \frac{1}{5}$ so $\frac{\text{Cov}(X,Z)}{\sqrt{1\times 25}} = \frac{1}{5}$, so $\text{Cov}(X,Z) = 1$.
Similarly $\rho_{Y,Z} = -\frac{1}{5}$ gives $\text{Cov}(X,Z) = -1$.
So
$$\text{Cov}(\mathbf{X}) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 25 \end{pmatrix}.$$

(b) We have
$$\begin{pmatrix} U \\ W \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} 0 \\ -4 \end{pmatrix} = A\mathbf{X} + \mathbf{b}$$
Then
$$E\begin{pmatrix} U \\ W \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -4 \\ 6 \end{pmatrix} + \begin{pmatrix} 0 \\ -4 \end{pmatrix} = \begin{pmatrix} -7 \\ 8 \end{pmatrix}$$
and
$$\text{Cov}\begin{pmatrix} U \\ W \end{pmatrix} = A\,\text{Cov}(\mathbf{X})A^T = \begin{pmatrix} 27 & -25 \\ -25 & 33 \end{pmatrix}.$$

(c) We observe $E((2X + Z - 6)^2) = E((W-2)^2)$. From (b) we have
$$\text{Var}(W-2) = \text{Var}(W) = 33$$
and $E(W) = 8$, so $E(W-2) = 6$. Hence, by rearranging $\text{Var}(W - 2) = E((W-2)^2) - (E(W-2))^2$,
$$E((W-2)^2) = \text{Var}(W-2) + (E(W-2))^2 = 33 + 6^2 = 69.$$

31. (a) The covariance matrix is the identity,

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(b) We have $\text{Cov}(R\mathbf{X}) = R\,\text{Cov}(\mathbf{X})R^T = RIR^T = RR^T$, and evaluating RR^T gives the identity (i.e. R is orthogonal), which was $\text{Cov}(X)$.

32. (a) We have

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)\text{var}(X_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{2}{1 \times \sqrt{5}} = 0.894.$$

X_1 and X_2 are positively correlated and they are not independent.

(b) We have

$$E(Y) = (1,2)E(\mathbf{X}) = (1,2)\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

$$\text{Cov}(Y) = (1,2)\,\text{Cov}(\mathbf{X})\begin{pmatrix} 1 \\ 2 \end{pmatrix} = (1,2)\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = (5,12)\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 29.$$

and

$$E(\mathbf{Z}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} E(\mathbf{X}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$$\text{Cov}(\mathbf{Z}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\text{Cov}(\mathbf{X})\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T$$
$$= \begin{pmatrix} 5 & 12 \\ 11 & 26 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 29 & 63 \\ 63 & 137 \end{pmatrix}.$$

Because the normal distribution is determined completely from its mean and its covariance matrix and because the transformations Y and $\mathbf{Z}$ are both linear, it follows that the distributions of Y and $\mathbf{Z}$ are

$$Y \sim N(0, 29) \quad \text{and} \quad \mathbf{Z} \sim N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 29 & 63 \\ 63 & 137 \end{pmatrix}\right].$$

33. (a) We have

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[x_1^2 - 2\rho x_1 x_2 + x_2^2\right]\right\},$$

so integrating out x_2 gives

$$\begin{aligned}
f_{X_1}(x_1) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\left[x_1^2 - 2\rho x_1 x_2 + x_2^2\right]}{2(1-\rho^2)}\right\} dx_2 \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\left[(x_2-\rho x_1)^2 - (\rho x_1)^2 + x_1^2\right]}{2(1-\rho^2)}\right\} dx_2 \\
&\quad \text{(by completing the square)} \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-x_1^2(1-\rho^2)}{2(1-\rho^2)}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right\} dx_2 \\
&\quad \text{(taking terms which don't involve } x_2 \text{ outside the integral)} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x_1^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{\frac{-(x_2-\rho x_1)^2}{2(1-\rho^2)}\right\} dx_2.
\end{aligned}$$

The term inside the integral here is the p.d.f. of a Normal random variable with mean ρx_1 and variance $(1-\rho^2)$. (Because the integral is with respect to x_2, we can treat x_1 as a constant, so this can make sense. Alternatively change variables to $y = x_2 - \rho x_1$, getting the p.d.f. of a Normal random variable with mean zero.) Hence the integral is 1, and so we have shown

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}},$$

which is the p.d.f. of the standard normal, as required.

$$\begin{aligned}
f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} \\
&= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[x_1^2-2\rho x_1x_2+x_2^2\right]\right\}}{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_1^2}{2}\right)} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{1}{2(1-\rho^2)}x_1^2+\frac{2\rho}{2(1-\rho^2)}x_1x_2-\frac{1}{2(1-\rho^2)}x_2^2+\frac{x_1^2}{2}\right] \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left\{-\frac{1}{2}\left[\frac{x_1^2-2\rho x_1x_2+x_2^2-(1-\rho^2)x_1^2}{1-\rho^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right]
\end{aligned}$$

and so $X_2|X_1 = x_1 \sim N(\rho x_1, 1-\rho^2)$ as required.

[Note the relationship between these calculations.]

34. Using the theory in the notes, $\mathbf{Y}$ has a bivariate normal distribution with mean vector

$$A\mu = \begin{pmatrix} \mu_1 - \frac{\mu_2\sigma_1}{\sigma_2} \\ \mu_2 + \frac{\mu_1\sigma_2}{\sigma_1} \end{pmatrix},$$

and covariance matrix

$$A\Sigma A^T = \begin{pmatrix} 1 & -\frac{\sigma_1}{\sigma_2} \\ \frac{\sigma_2}{\sigma_1} & 1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & \frac{\sigma_2}{\sigma_1} \\ -\frac{\sigma_1}{\sigma_2} & 1 \end{pmatrix} = \begin{pmatrix} \frac{2\sigma_1(\sigma_1\sigma_2-\sigma_{12})}{\sigma_2} & 0 \\ 0 & \frac{2\sigma_2(\sigma_1\sigma_2+\sigma_{12})}{\sigma_1} \end{pmatrix}.$$

So

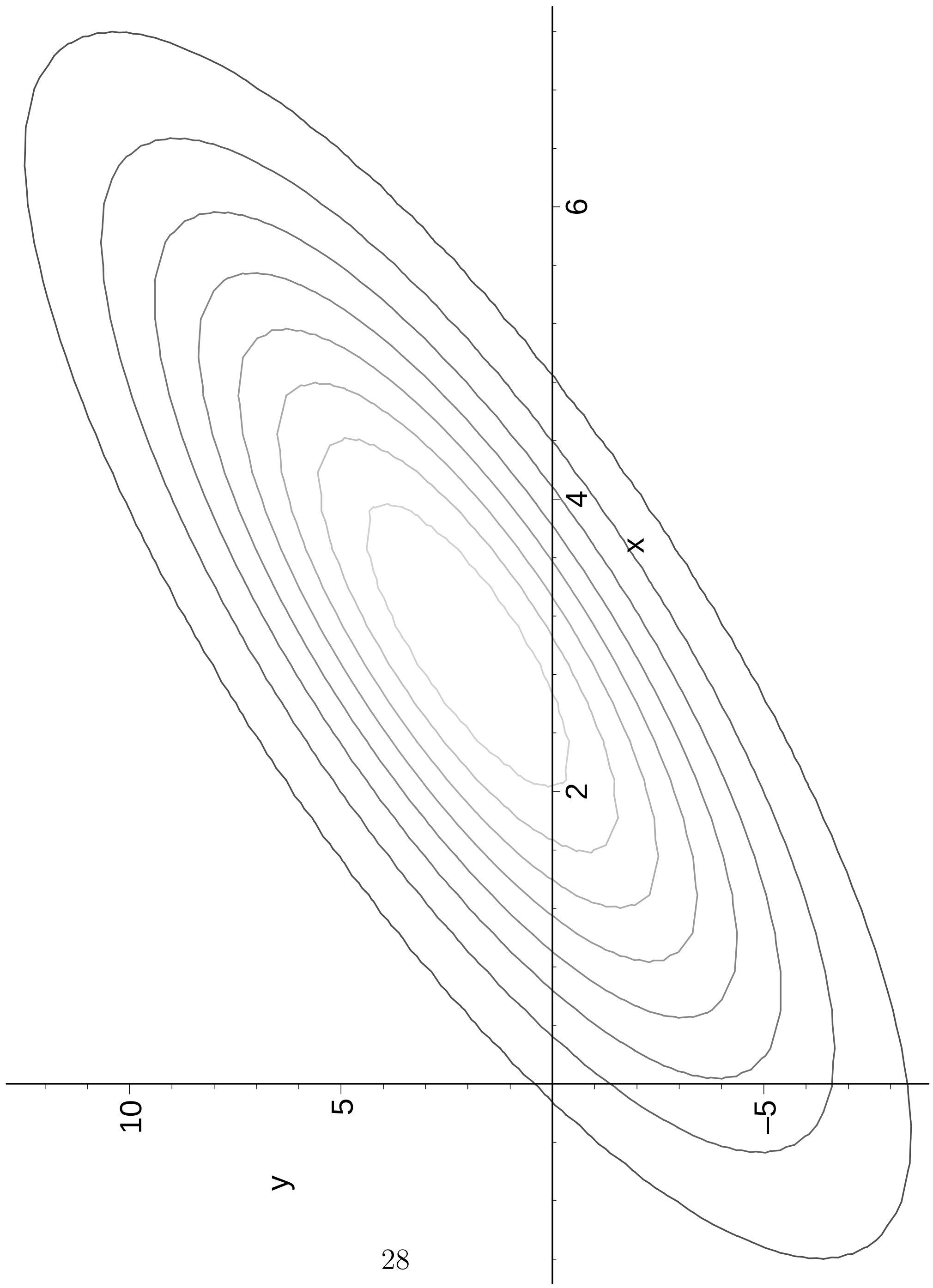
$$\mathbf{Y} \sim N_2\left(\begin{pmatrix} \mu_1 - \frac{\mu_2\sigma_1}{\sigma_2} \\ \mu_2 + \frac{\mu_1\sigma_2}{\sigma_1} \end{pmatrix}, \begin{pmatrix} \frac{2\sigma_1(\sigma_1\sigma_2-\sigma_{12})}{\sigma_2} & 0 \\ 0 & \frac{2\sigma_2(\sigma_1\sigma_2+\sigma_{12})}{\sigma_1} \end{pmatrix}\right),$$

and so the components of $\mathbf{Y}$ are independent. This transformation can be applied to any bivariate normal random vector, so any bivariate normal random vector can be transformed by a linear transformation into a vector of independent normal random variables. (There are other transformations which do this; see also the section in the notes on conditional distributions.)

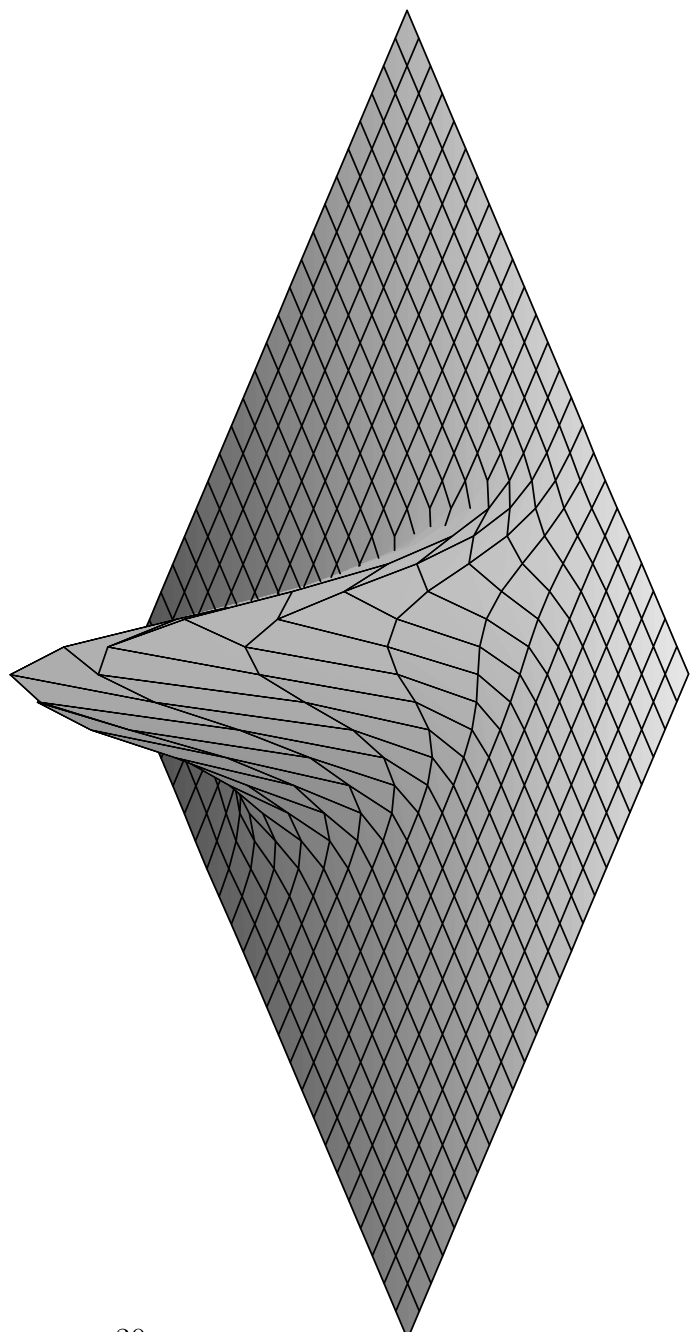
35. The contour plot of the p.d.f. of

$$N_2\left(\begin{pmatrix}3\\2\end{pmatrix}, \begin{pmatrix}4 & 8\\8 & 25\end{pmatrix}\right)$$

looks like


y
10
5
–5
x
2
4
6

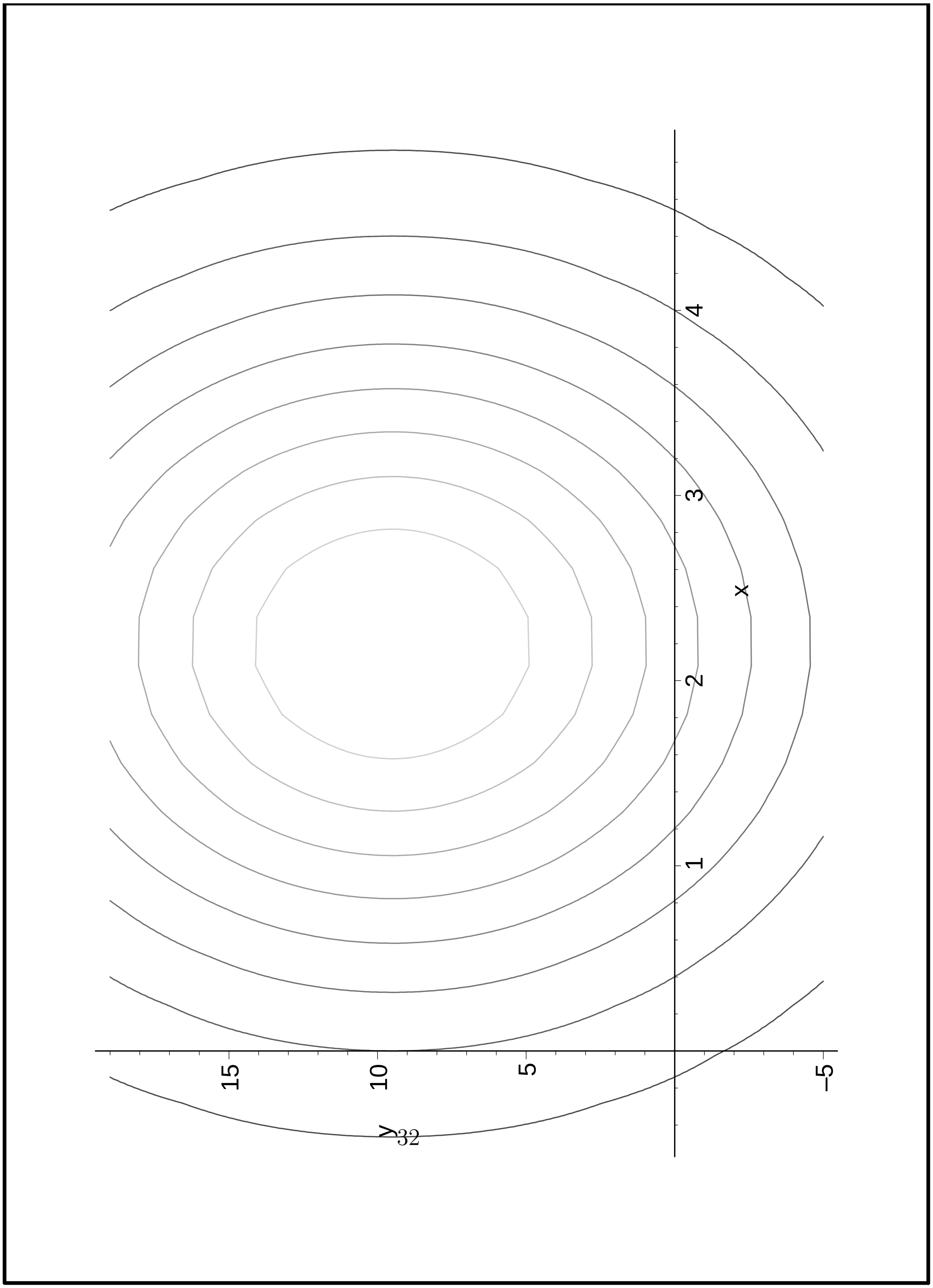
and the 3D plot looks like

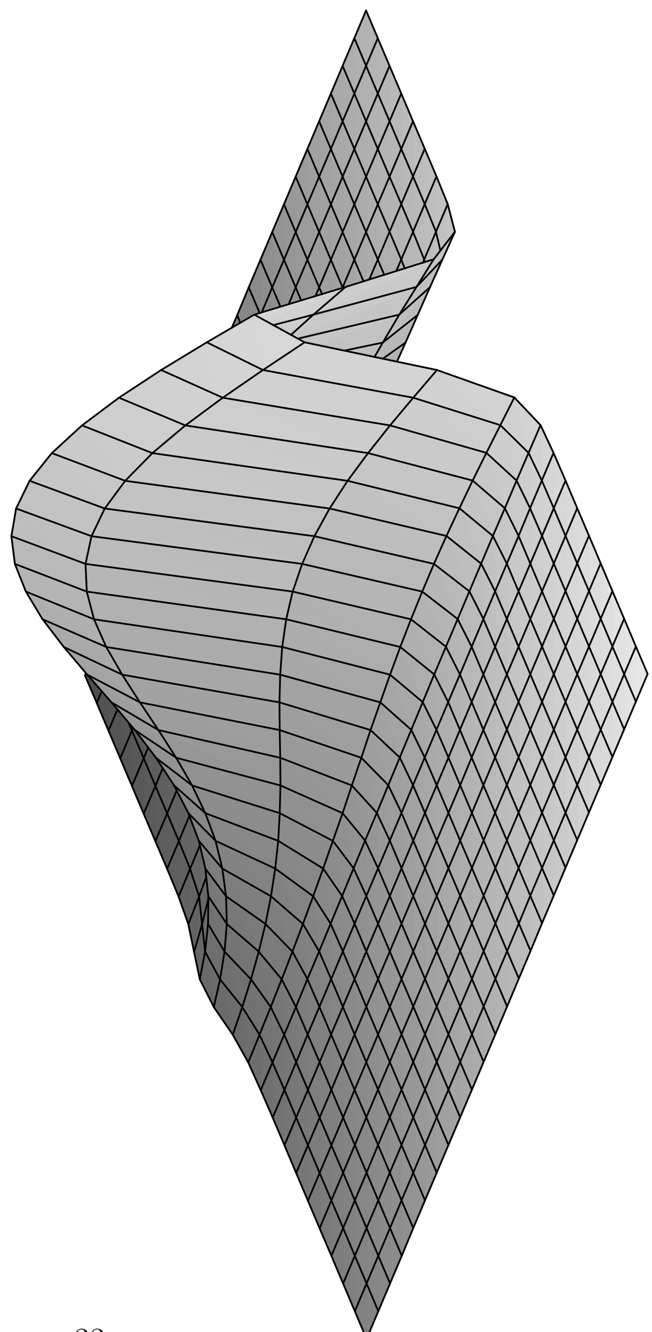


Applying the transformation described in Question 34 gives the distribution

$$N_2\left(\begin{pmatrix}\frac{11}{5}\\ \frac{19}{2}\end{pmatrix}, \begin{pmatrix}\frac{8}{5} & 0\\ 0 & 90\end{pmatrix}\right),$$

giving the plots:


x
1
2
3
4
y
15
10
5
–5



For the independent case (after the transformation) the axes of the ellipses are parallel to the axes, whereas for the initial distribution (where the two variables have a strong positive correlation) the ellipse is quite eccentric with the axes diagonally oriented; large values of X_1 tend to correspond to large values of X_2. [Depending on the choice of limits for the plot, the contour plot of the independent case may appear circular; this is an effect caused by Maple choosing different scales for the axes.]

36. (a) From the covariance matrix, the correlation coefficient between X_1 and X_2 is $\frac{-30}{\sqrt{144}\sqrt{25}} = -\frac{1}{2}$, that between X_1 and X_3 is $\frac{48}{\sqrt{144}\sqrt{64}} = \frac{1}{2}$ and that between X_2 and X_3 is $\frac{10}{\sqrt{25}\sqrt{64}} = \frac{1}{4}$.

    (b) We have $\mathbf{Y} = A\mathbf{X}$ where

$$A = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}.$$

So the mean vector of $\mathbf{Y}$ is $A\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, and its covariance matrix is

$$A\Sigma A^T = \begin{pmatrix} 304 & -212 \\ -212 & 229 \end{pmatrix},$$

and so the correlation between Y_1 and Y_2 is $\frac{-212}{\sqrt{304}\sqrt{229}} = -0.803\ldots$.

37. (a) We note that $\mathbf{X} \sim N_n(\mathbf{0}, \sigma^2 I)$, where $\mathbf{0}$ is an n-dimensional vector of zeros and I is the n-dimensional identity matrix. The mean vector of $\mathbf{Y}$ will then be $R\mathbf{0} = \mathbf{0}$, and the covariance matrix will be $R(\sigma^2 I)R^T = \sigma^2 RR^T = \sigma^2 I$, by the orthogonality of R, so the multivariate normal theory tells us that $\mathbf{Y}$ also has $N_n(\mathbf{0}, \sigma^2 I)$ distribution.

(b) By the form given for R,

$$Y_1 = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} X_i = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} X_i = \sqrt{n}\bar{X}.$$

By the orthogonality of R, $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2$, so

$$\begin{aligned} \sum_{i=2}^{n} Y_i^2 &= \sum_{i=1}^{n} X_i^2 - Y_1^2 \\ &= \sum_{i=1}^{n} X_i^2 - n(\bar{X})^2. \end{aligned}$$

(c) By part (a), the Y_i are independent $N(0, \sigma^2)$ random variables, so $\frac{Y_i}{\sigma} \sim N(0,1)$, and by Exercise 29

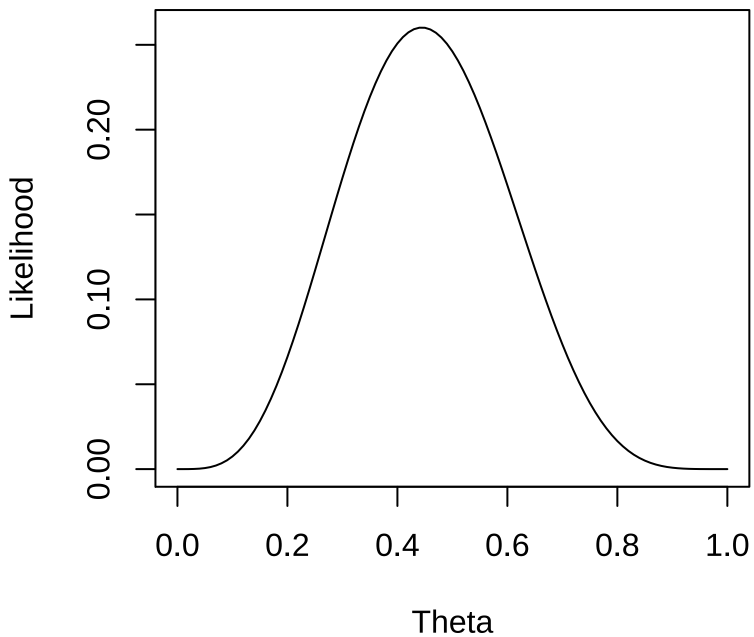
$$\sum_{i=2}^{n} \left(\frac{Y_i}{\sigma}\right)^2 \sim \chi^2_{n-1},$$

and hence by part (b) we have

$$\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \sim \chi^2_{n-1},$$

as required. The independence follows because Y_1 is independent of the remaining variables by the form of the covariance matrix and multivariate normal theory.

38. (a) The likelihood will be $L(\theta; 4) = \binom{9}{4}\theta^4(1-\theta)^5 = 126\theta^4(1-\theta)^5$. The following is a plot of this from `R`.



(b) The likelihood will now be

$$L(n;4) = \binom{n}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^{n-4} = \binom{n}{4}\frac{81}{4^n}.$$

The values requested are given in the table:

	$L(n;4)$
$n=4$	$\frac{81}{256} = 0.316$
$n=5$	$\frac{405}{1024} = 0.396$
$n=6$	$\frac{1215}{4096} = 0.297$
$n=7$	$\frac{2835}{16384} = 0.173$

39. The likelihood function is $L(\theta;\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$.

(a) The contribution to the likelihood from observation i will be $f(x_i|\theta) = \frac{1}{\theta}e^{-\frac{1}{\theta}x_i}$. So the likelihood is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \frac{1}{\theta}e^{-\frac{1}{\theta}x_i} = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}.$$

The parameter set $\Theta = (0,\infty)$.

(b) The contribution to the likelihood from observation i will be $f(x_i|\theta) = \binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i}$, for $x_i = 0,1,\ldots,m$. So the likelihood is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i} = \left[\prod_{i=1}^n \binom{m}{x_i}\right]\theta^{\sum_{i=1}^n x_i}(1-\theta)^{nm-\sum_{i=1}^n x_i}.$$

The parameter set $\Theta = [0,1]$.

(c) The contribution to the likelihood from observation i will be $f(x_i|\theta) = \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x_i-\mu)^2}{2\theta}}$. So the likelihood is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x_i-\mu)^2}{2\theta}} = \frac{1}{(2\pi\theta)^{n/2}}e^{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2}.$$

The parameter set $\Theta = (0,\infty)$.

40. (a) The contribution to the likelihood from observation i will be $f(x_i|\theta) = \frac{4^\theta}{\Gamma(\theta)}x_i^{\theta-1}e^{-4x_i}$. So the likelihood is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \frac{4^\theta}{\Gamma(\theta)}x_i^{\theta-1}e^{-4x_i} = \frac{4^{n\theta}}{\{\Gamma(\theta)\}^n}\left(\prod_{i=1}^n x_i\right)^{\theta-1} e^{-4\sum_{i=1}^n x_i}.$$

The parameter set $\Theta = (0,\infty)$.

(b) The p.d.f. of this distribution can be found using the theory in section 1.6. If $U \sim Ga(\alpha, \beta)$ then we want the p.d.f. of $X = g(U) = 1/U$. The range of U is $\mathbb{R}^+$ so g can be considered as a decreasing function, with inverse $g^{-1}(x) = 1/x$, and $f_U(u) = \frac{\beta^\alpha u^{\alpha-1} e^{-\beta u}}{\Gamma(\alpha)}$ for $u > 0$, so by the method for monotonic functions in section 1.6 we have

$$f_X(x) = \frac{\beta^\alpha x^{1-\alpha} e^{-\beta/x}}{\Gamma(\alpha)} \left| -\frac{1}{x^2} \right| = \frac{\beta^\alpha x^{-(1+\alpha)} e^{-\beta/x}}{\Gamma(\alpha)},$$

for $x > 0$ (and zero otherwise).

So, with $\alpha = 1$ and $\beta = \theta$, the contribution of observation i to the likelihood is $f(x_i|\theta) = \theta x_i^{-2} e^{-\theta/x_i}$, for $x_i > 0$. So the likelihood is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \theta x_i^{-2} e^{-\theta/x_i} = \theta^n \left(\prod_{i=1}^n x_i \right)^{-2} e^{-\theta \sum_{i=1}^n x_i^{-1}}.$$

The parameter set $\Theta = (0, \infty)$.

(c) The contribution to the likelihood from observation i will be $f(x_i|\theta) = \frac{1}{B(\theta,\theta)} x_i^{\theta-1}(1 - x_i)^{\theta-1}$. So the likelihood is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{B(\theta, \theta)} x_i^{\theta-1} (1-x_i)^{\theta-1} = \frac{1}{[B(\theta, \theta)]^n} \left(\prod_{i=1}^n x_i \right)^{\theta-1} \left[\prod_{i=1}^n (1 - x_i) \right]^{\theta-1}.$$

The parameter set $\Theta = (0, \infty)$.

41. (a) The log likelihood is

$$\log \left(\frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^n x_i} \right) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^n x_i.$$

(b) The log likelihood is

$$\log \left(\left[\prod_{i=1}^n \binom{m}{x_i} \right] \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{nm - \sum_{i=1}^n x_i} \right)$$

which is

$$\sum_{i=1}^{n} \log \binom{m}{x_i} + \sum_{i=1}^{n} x_i \log\theta + (nm - \sum_{i=1}^{n} x_i)\log(1-\theta).$$

(c) The log likelihood is

$$\log\left(\frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2}\right) = -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2.$$

42. The probability of observing the values 4, 0 and 3 as independent observations from a $Po(\lambda)$ population is

$$\frac{\lambda^4 e^{-\lambda}}{4!}\frac{e^{-\lambda}}{0!}\frac{\lambda^3 e^{-\lambda}}{3!} = \frac{\lambda^7 e^{-3\lambda}}{144},$$

which gives 3.46×10^{-4} for $\lambda = 1$, 2.20×10^{-3} for $\lambda = 2$, and 1.87×10^{-3} for $\lambda = 3$. So the maximum likelihood estimate is $\lambda = 2$.

43. (a) From question 38(a), the likelihood is $L(\theta; 4) = 126\theta^4(1-\theta)^5$. Differentiating,

$$\frac{dL}{d\theta} = 126\theta^3(1-\theta)^4(4(1-\theta) - 5\theta).$$

This will be zero when $\theta = 0$, $\theta = 1$ or when $4(1-\theta) = 5\theta$, i.e. when $\theta = \frac{4}{9}$. To check for maxima, calculate the second derivative:

$$\frac{d^2L}{d\theta^2} = 126(12\theta^2(1-\theta)^5 - 40\theta^3(1-\theta)^4 + 20\theta^4(1-\theta)^3).$$

At $\theta = \frac{4}{9}$ this is negative, so $\theta = \frac{4}{9}$ is a local maximum; at the other two it is zero, which is inconclusive, but checking the plot and the values of the likelihood confirms that $\theta = \frac{4}{9}$ is the maximum.

(b) From question 38(b), the likelihood is $L(n;4) = \binom{n}{4}\frac{81}{4^n}$, and from the values for $n = 4, 5, 6, 7$ calculated in that question the highest value was for $n = 5$. To confirm that $n = 5$ gives the highest value of the likelihood for integers $n \geq 4$, we can check the ratio

$$\frac{L(n+1;4)}{L(n;4)} = \frac{n+1}{4(n-3)},$$

which is less than 1 if $n \geq 5$. Hence for $n \geq 5$ the likelihood is decreasing in n, so 5 is the maximum likelihood estimate of n.

44. From question 39(b), the likelihood function is

$$L(\theta;\mathbf{x}) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{\sum_{i=1}^n (m-x_i)} \prod_{i=1}^n C_{x_i}^m$$

and the log-likelihood is

$$\ell(\theta;\mathbf{x}) = \sum_{i=1}^n x_i \log\theta + \sum_{i=1}^n (m-x_i)\log(1-\theta) + \sum_{i=1}^n \log C_{x_i}^m.$$

The first derivative of $\ell(\theta;\mathbf{x})$ is

$$\frac{d\ell(\theta;\mathbf{x})}{d\theta} = \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{1-\theta}\sum_{i=1}^n (m-x_i)$$

and equalizing this to zero we get

$$\frac{d\ell(\theta;\mathbf{x})}{d\theta} = 0 \Rightarrow \frac{1}{\theta}\sum_{i=1}^n x_i = \frac{1}{1-\theta}\sum_{i=1}^n (m-x_i)$$

$$\theta nm - \theta\sum_{i=1}^n x_i = \sum_{i=1}^n x_i - \theta\sum_{i=1}^n x_i \Rightarrow \theta nm = \sum_{i=1}^n x_i \Rightarrow \theta = \frac{1}{nm}\sum_{i=1}^n x_i = \hat{\theta}.$$

The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-\theta)^2}\sum_{i=1}^{n}(m - x_i) < 0$$

for all θ. Hence $\hat{\theta}$ is the required maximum likelihood estimate.

45. From question 39(c), the likelihood function is

$$L(\theta; \mathbf{x}) = \frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}\sum(x_i-\mu)^2}$$

and so the log-likelihood function is

$$\ell(\theta; \mathbf{x}) = -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2 = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2.$$

The first derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

and setting this equal to zero

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = 0 \Rightarrow \frac{n}{2\theta} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\theta^2} \Rightarrow \theta = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \hat{\theta}.$$

The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{\theta^2}\left(\frac{n}{2} - \frac{1}{\theta}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

and evaluating it at $\theta = \hat{\theta}$ gives

$$\left.\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2}\right|_{\theta=\hat{\theta}} = \frac{1}{\hat{\theta}^2}\left(\frac{n}{2} - \frac{n}{\sum_{i=1}^{n}(x_i - \mu)^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) = -\frac{n}{2\hat{\theta}^2} < 0.$$

Hence $\hat{\theta}$ is the required maximum likelihood estimate.

For the last part, $\hat{\theta}$ is unbiased because

$$E(\hat{\theta}) = \frac{1}{n} E \sum_{i=1}^{n} (X_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2.$$

Note: this is the usual estimator of the variance of a normal population with **known** mean. (The factor of $n-1$ in the usual formula for the sample variance assumes the mean is unknown.)

46. The likelihood function is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} \frac{\theta^3}{\Gamma(3)} x_i^{3-1} e^{-\theta x_i} = \frac{\theta^{3n}}{2^n} \left(\prod_{i=1}^{n} x_i \right)^2 e^{-\theta \sum_{i=1}^{n} x_i}$$

and so the log-likelihood function is

$$\ell(\theta; \mathbf{x}) = 3n \log \theta - n \log 2 + 2 \sum_{i=1}^{n} \log x_i - \theta \sum_{i=1}^{n} x_i$$

The first derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = \frac{3n}{\theta} - \sum_{i=1}^{n} x_i$$

and setting this equal to zero we get

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = 0 \Rightarrow \frac{3n}{\theta} = \sum_{i=1}^{n} x_i \Rightarrow \theta = \frac{3n}{\sum_{i=1}^{n} x_i} = \hat{\theta}.$$

The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{3n}{\theta^2},$$

which is negative for all θ and so $\hat{\theta}$ is the required maximum likelihood estimate.

From $\bar{x} = 3$, we get $\sum_{i=1}^{n} x_i = n\bar{x} = 3n$ and so $\hat{\theta} = 3n/(3n) = 1$.

47. The log-likelihood function is

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = \log \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) = -(\alpha+1)\sum_{i=1}^{n} \log x_i + n\log\alpha + n\alpha\log\beta,$$

assuming $\alpha > 0$, $\beta > 0$ and $\beta \leq x_i$ for all i. (If $\beta > x_i$ for any i then the likelihood is zero.) The first partial derivatives are

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \alpha} = -\sum_{i=1}^{n} \log x_i + \frac{n}{\alpha} + n\log\beta \quad (4)$$

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \beta} = n\alpha/\beta \quad (5)$$

(for $\alpha, \beta > 0$ and $\beta \leq x_i$ for all i).

Hence $\ell(\boldsymbol{\theta}; \mathbf{x})$ is increasing in β for $\beta \leq x_i$ for all i and all $\alpha > 0$, so to maximise β we take $\hat{\beta} = \min(x_1, \ldots, x_n)$. Then from (4) we have that

$$\frac{\partial \ell(\hat{\boldsymbol{\theta}}; \mathbf{x})}{\partial \hat{\alpha}} = 0$$

implies that

$$\frac{n}{\hat{\alpha}} = \sum_{i=1}^{n} \log x_i - n\log\hat{\beta}$$

and so there is a possible maximum at

$$\hat{\alpha} = \frac{1}{\log\left(\prod_{i=1}^{n} x_i^{1/n} / \min(x_1, \ldots, x_n)\right)}.$$

Checking the second derivative,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \alpha^2} = -\frac{n}{\alpha^2} < 0,$$

for all $\alpha > 0$ and for all $\beta \leq x_i$.

So $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^T$ is the required maximum likelihood estimate of $\boldsymbol{\theta}$.

48. The log-likelihood function is

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = \log \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \log x_i - n \log C(\alpha, \beta),$$

for $\alpha \leq x_i \leq \beta$, where $C(\alpha, \beta) = \beta(\log \beta - 1) - \alpha(\log \alpha - 1)$.

The function C is decreasing in α (for fixed β) and increasing in β (for fixed α). So ℓ is increasing in α (for fixed β) and it is decreasing in β (for fixed α). To maximize ℓ we need to take the largest possible value of α, which is $\hat{\alpha} = \min(x_1, \ldots, x_n)$ (since $\alpha \leq x_i$) and the smallest possible value for β, which is $\hat{\beta} = \max(x_1, \ldots, x_n)$ (since $\beta \geq x_i$).

So $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^T$ is the maximum likelihood estimate of $\boldsymbol{\theta}$.

49. The log-likelihood function is

$$\ell(\theta; \mathbf{x}) = \log \prod_{i=1}^{n} p(x_i) = \log(1-\theta) \sum_{i=1}^{n} x_i + n \log \theta$$

The first derivative of ℓ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = -\frac{\sum_{i=1}^{n} x_i}{1-\theta} + \frac{n}{\theta}$$

and this is zero if and only if

$$\theta = \hat{\theta} = \frac{n}{n + \sum_{i=1}^{n} x_i}$$

The second derivative of ℓ is

$$\frac{d^2\ell(\theta;\mathbf{x})}{d\theta^2} = -\frac{\sum_{i=1}^n x_i}{(1-\theta)^2} - \frac{n}{\theta^2} < 0,$$

for all θ (because $x_i > 0$) and so $\hat{\theta}$ is the required maximum likelihood estimate of θ.

50. (a) The log-likelihood function is

$$\ell(\theta;\mathbf{x}) = \frac{n}{2}\log\theta - \frac{1}{2}\sum_{i=1}^n \log(2\pi x_i^3) - \frac{\theta}{2\mu^2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{x_i}$$

The first derivative of ℓ is

$$\frac{d\ell(\theta;\mathbf{x})}{d\theta} = \frac{n}{2\theta} - \frac{1}{2\mu^2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{x_i}$$

and for this to be zero we must have

$$\theta = \hat{\theta} = \frac{n\mu^2}{\sum_{i=1}^n (x_i-\mu)^2 x_i^{-1}}.$$

The second derivative of ℓ is

$$\frac{d^2\ell(\theta;\mathbf{x})}{d\theta^2} = -\frac{n}{2\theta^2} < 0,$$

for all θ and so $\hat{\theta}$ is the required maximum likelihood estimate of θ.

(b) The log-likelihood function is

$$\ell(\mu,\theta;\mathbf{x}) = \frac{n}{2}\log\theta - \frac{1}{2}\sum_{i=1}^n \log(2\pi x_i^3) - \frac{\theta}{2\mu^2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{x_i}.$$

We now calculate the first partial derivatives. Differentiating with respect to μ and rearranging gives

$$\begin{aligned}\frac{\partial \ell(\mu,\theta;\mathbf{x})}{\partial \mu} &= \frac{\theta}{\mu^3}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}+\frac{\theta}{\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)}{x_i}\\ &= \sum_{i=1}^{n}\frac{\theta}{\mu^3}\frac{x_i-\mu}{x_i}\left((x_i-\mu)+\mu\right),\end{aligned}$$

giving

$$\frac{\partial \ell(\mu,\theta;\mathbf{x})}{\partial \mu} = \frac{\theta}{\mu^3}\sum_{i=1}^{n}(x_i-\mu). \tag{6}$$

Differentiating with respect to θ gives

$$\frac{\partial \ell(\mu,\theta;\mathbf{x})}{\partial \theta} = \frac{n}{2\theta}-\frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}. \tag{7}$$

For these to both be zero, from (6) we must have

$$\mu=\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i,$$

and as before we must have

$$\theta=\frac{n\mu^2}{\sum_{i=1}^{n}(x_i-\mu)^2x_i^{-1}},$$

so let

$$\hat{\theta}=\frac{n\hat{\mu}^2}{\sum_{i=1}^{n}(x_i-\hat{\mu})^2x_i^{-1}}.$$

Then both first partial derivatives will be zero at $(\hat{\mu},\hat{\theta})$. The second partial derivatives are

$$\begin{aligned}
\frac{\partial^2 \ell(\mu,\theta;\mathbf{x})}{\partial \mu^2} &= -\frac{3\theta}{\mu^4}\sum_{i=1}^n (x_i-\mu) - n\frac{\theta}{\mu^3}, && (8)\\
\frac{\partial^2 \ell(\mu,\theta;\mathbf{x})}{\partial \mu \partial\theta} &= \frac{1}{\mu^3}\sum_{i=1}^n (x_i-\mu), && (9)\\
\frac{\partial^2 \ell(\mu,\theta;\mathbf{x})}{\partial \theta^2} &= -\frac{n}{2\theta^2}. && (10)
\end{aligned}$$

So the Hessian evaluated at $(\hat{\mu},\hat{\theta})$ is

$$\begin{pmatrix} -\frac{n\hat{\theta}}{\hat{\mu}^3} & 0 \\ 0 & -\frac{n}{2\hat{\theta}^2} \end{pmatrix},$$

which is negative definite and so $(\hat{\mu},\hat{\theta})$ is a maximum. Hence $(\hat{\mu},\hat{\theta})$ is the MLE of (μ,θ).

51. (a) i. The log likelihood is $\log\binom{9}{4} + 4\log\theta + 5\log(1-\theta)$, and the MLE is $4/9$. So we want to solve

$$4\log\theta + 5\log(1-\theta) \geq 4\log\frac{4}{9} + 5\log(1-\frac{4}{9}) - 2.$$

Solving this in Maple gives the range of values $[0.161, 0.757]$.

ii. The approximate confidence interval calculation gives $[0.120, 0.769]$. This is similar, but a bit wider, especially at the lower end.

(b) The range of values with likelihoods within 2 of the maximum is now $[0.342, 0.550]$. The approximate confidence interval gives $[0.342, 0.547]$, which is very similar (slightly narrower at the top end).

52. (a) Let $g(x) = 4/x$ so that $Y = g(X)$; then $g^{-1}(y) = 4/y$ and $\frac{d}{dy}g^{-1}(y) = -4/y^2$. So

$$f_Y(y) = f_X(4/y)\left|-\frac{4}{y^2}\right| = 12\frac{4}{y}\left(1-\frac{4}{y}\right)^2\frac{4}{y^2} = 192\frac{(y-4)^2}{y^5}$$

and, since $0 < x < 1$, the range of y is given by $y > 4$.

(b) We have

$$\begin{aligned} E\left(\frac{1}{X}\right) &= \int_{-\infty}^{\infty} \frac{1}{x} f_X(x)\,dx = \int_0^1 12\frac{1}{x}x(1-x)^2\,dx = 12\int_0^1 (1-x)^2\,dx \\ &= 12\int_0^1 (1-2x+x^2)\,dx \\ &= 12[x - x^2 + x^3/3]_0^1 = 12(1-1+1/3) = 4. \end{aligned}$$

Since $Y = 4/X$, we have

$$E(Y) = 4E\left(\frac{1}{X}\right) = 4 \times 4 = 16.$$

53. (a) From the correlations,

$$\begin{aligned} \lambda_1 &= \mathrm{Cov}(X_1, X_2) = 0.4 \times 1 \times \sqrt{k} = 0.4\sqrt{k} \\ \nu_1 &= \mathrm{Cov}(X_1, X_3) = 0 \end{aligned}$$

To find k, we have

$$0.9 = \rho_{X_2,X_3} = \frac{\mathrm{Cov}(X_2, X_3)}{\sqrt{\mathrm{Var}(X_2)\,\mathrm{Var}(X_3)}} = \frac{1}{\sqrt{2k}},$$

so $k = 0.62$.

Then $\lambda_1 = 0.4 \times \sqrt{0.62} = 0.31$. Because V needs to be symmetric, we have $\lambda_2 = \lambda_1 = 0.31$, $\nu_2 = \nu_1 = 0$ and $k = 0.62$. So we have

$$V = \begin{pmatrix} 1 & 0.31 & 0 \\ 0.31 & 0.62 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

(b)

$$(X_2, X_3)^T \sim N_2\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0.62 & 1 \\ 1 & 2 \end{pmatrix}\right]$$

(a bivariate normal distribution).

(c) We have

$$Y = \begin{pmatrix} 2 & -1 & 1/2 \end{pmatrix} \mathbf{X} = A\mathbf{X}.$$

So

$$E(Y) = AE(\mathbf{X}) = \begin{pmatrix} 2 & -1 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = 2$$

and

$$\mathrm{Var}(Y) = AVA^T = \begin{pmatrix} 2 & -1 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 0.31 & 0 \\ 0.31 & 0.62 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 1/2 \end{pmatrix} = 2.88.$$

Since Y is a linear transformation of $\mathbf{X}$, which follows a normal distribution, the distribution of Y will also be normal and so we have

$$Y \sim N(2, 2.88).$$

54. We have $E(Y) = E(2X + Z) = 2E(X) + E(Z) = 2 \times 1 + 0 = 2$. Also

$$E\{E(Y|X)\} = E(2X) + E(Z) = 2 \times 1 = 2 = E(Y).$$

Also $\mathrm{Var}(Y) = 4\,\mathrm{Var}(X) + \mathrm{Var}(Z) = 4 \times 2 + 1 = 9$. And

$$E\{\mathrm{Var}(Y|X)\} + \mathrm{Var}\{E(Y|X)\} = E(1) + \mathrm{Var}(2X) = 1 + 2 \times 2 = 9 = \mathrm{Var}(Y).$$

When we work out $\mathrm{Var}(Y)$ directly from the model $Y = X + Z$, we assume that X and Z are independent and so their covariance is zero. When we work out $\mathrm{Var}(Y)$ using the conditional expectations and variances, we assume that X and Z are independent and so $\mathrm{Var}(Z|X) = \mathrm{Var}(Z) = 1$.

55. The likelihood function is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} p(x_i) = \theta^{nr} \prod_{i=1}^{n} \begin{pmatrix} x_i + r - 1 \\ r - 1 \end{pmatrix} (1 - \theta)^{x_i}.$$

The log-likelihood function is

$$\ell(\theta; \mathbf{x}) = (nr)\log\theta + \sum_{i=1}^{n}\log\left(\begin{array}{c} x_i + r - 1 \\ r - 1 \end{array}\right) + \log(1-\theta)\sum_{i=1}^{n} x_i.$$

The first derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = \frac{nr}{\theta} - \frac{\sum_{i=1}^{n} x_i}{1-\theta}$$

and setting this equal to zero we get

$$\frac{nr}{\theta} = \frac{\sum_{i=1}^{n} x_i}{1-\theta}$$

which implies that

$$\theta = \widehat{\theta} = \frac{nr}{nr + \sum_{i=1}^{n} x_i}.$$

The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{nr}{\theta^2} - \frac{\sum_{i=1}^{n} x_i}{(1-\theta)^2},$$

which is negative, for all θ. So $\widehat{\theta}$ is the maximum likelihood estimate of θ.